Constructive Biology and Approaches to Temporal Grounding in Post-Reactive Robotics

Chrystopher L. Nehaniv[a], Kerstin Dautenhahn[b,c], and Martin J. Loomes[a]

[a]Faculty of Engineering and Information Sciences, University of Hertfordshire
College Lane, Hatfield Herts AL10 9AB, United Kingdom

[b]Department of Cybernetics, University of Reading
Whiteknights, P.O. Box 225, Reading RG6 6AY, United Kingdom

[c]MIT Artificial Intelligence Laboratory, 545 Technology Square, Cambridge MA 02139, U.S.A.

ABSTRACT

Constructive Biology (as opposed to descriptive biology) means understanding biological mechanisms through building systems that exhibit life-like properties. Applications include learning engineering tricks from biological systems, as well as the validation in biological modelling. In particular, biological systems (unlike reactive robots) in the course of development and experience become temporally grounded. Researchers attempting to transcend mere reactivity have been inspired by the drives, motivations, homeostasis, hormonal control, and emotions of animals. In order to contextualize and modulate behavior, these ideas have been introduced into robotics and synthetic agents, while further flexibility is achieved by introducing learning.

Broadening scope of the temporal horizon further requires post-reactive techniques that address not only the action in the *now*, although such action may perhaps be modulated by drives and affect. Support is needed for expressing and benefitting from past experiences, predictions of the future, and from interaction histories of the self with the world and with other agents (*stories*). Mathematical methods provide a new way to support such grounding in the construction of post-reactive systems. Moreover, the communication of such mathematical encoded histories of experience between situated agents opens a route to narrative intelligence, analogous to communication or story-telling in (biological) societies.

1. INTRODUCTION

1.1. Reactive Control

Reactivity is the control of action based immediately on stimuli present in the surrounding environment with minimal use of state internal information. This works well for very simple behaviors such as wall-following. It is even appears to be of primary importance in many living systems. But for more complex behaviors a wider temporal scope in order to better contextualize actions is needed. We address problems, grounding, and methods of post-reactive robotics that are intended to deal with such issues from a viewpoint of constructive biology. Instead of traditional AI methods such as path-planning, exhaustive mapping, and symbolic reasoning with distilled representations, more biologically realistic and computationally feasible approaches are inspired by the study of animals systems and interaction.

1.2. Constructive Biology

Constructive Biology is biology motivated by the desire to understand how biological systems actually are constructed by nature and develop over time rather than just to obtain a descriptive understanding. Conversely, it is also concerned with learning engineering "tricks" from biology that can be applied in artificial systems such as navigation systems, robotics, and artificial agents. The viewpoint is that one's understanding should enable one to, in principle, *build* the systems of interest.

For instance, John Von Neumann, in the late 1940s at a time when the biological mechanisms of genetic inheritance were still a mystery, sought to understand how living systems could reproduce. He succeeded with the help of Stanislaw Ulam and Arthur Burks in showing a particular nontrivial (cellular) automata model could indeed produce

---

Corresponding Author: C. L. Nehaniv,
E-mail: `C.L.Nehaniv@herts.ac.uk`, WWW: `http://www.cs.herts.ac.uk/~nehaniv`

a copy of itself.[1] His self-reproducing system used the same (inherited) information in two ways: (1) as unanalyzed data that are blindly copied, and (2) as in instructions to be executed in the construction of the replicant daughter automaton.[2] Later following the illumination of the structure of DNA by Watson and Crick,[3] we know that biological molecules of genetic inheritance have these same properties. In a more recent example, Barbara Webb has shown through building that a much simpler mechanism than expected, not involving functional decomposition or planning, is sufficient to account for much observed cricket phonotaxis behavior.[4,5] Valentino Braitenberg's thought experiments about simple robots[6] to which human observers attribute such states as 'fear', 'aggression', 'love', etc., illustrate that meaning of an interaction for an external observer can be quite different than it has for the agent (in these cases, simple taxis). The constructive biology of multi-agent systems will inescapably lead to mappings between channels of meaning that respect structural constraints and grounding of agents.

1.3. First and Second Person

Experiences directly of the robot, animal or other agent itself are called "first person". The experience of another is "second person".[7] Attribution of intentions and understanding of the behavior of a second person proceeds via some or all of:
(0) recognition of the other as a potential interaction partner (rather than a non-agentive object);
(1) the recognition of the another agent as similar to the first person partner;
(2) the projection or mapping of first person experience of *meaning* (see sec. 3.3 below) in the present (current state) to the second, making its actions, displays and signals (behavior) interpretable;
(3) the projection or mapping of first person past or temporally removed experience (narrative, stories, or history) to the second, attributing to the second person a historical groundedness and biography.
Cases of (2) are called *empathic resonance*, and are situated in the *now*. Cases of (3) involve a wider temporal horizon – extending from the past toward the now and from the imminent future toward the now. These cases are called *biographic reconstruction*.[8,9] This is how meaning between agents is transferred socially. Agents capable of dynamically reconstructing the biographies (histories) of the self and/or others during their life times are called *autobiographic agents*.

In contrast to behaviorists like Skinner,[10] a constructive biologist need not be restricted to external observation of stimuli and responses, rejecting speculation of what occurs inside the organism. Indeed, the engineering and design of internal mechanisms are just aspects of the experimental and theoretical apparatus that the constructive biologist may manipulate, vary, and control. Just as with studying and building improved sensors and actuators in artificial agents and robots, mechanisms of internal control, remembering, predicting, and possibly empathizing and biographical reconstruction for second persons can be the object of scientific inquiry.

The study of correspondence via the algebraic notion of homomorphism (full, partial or relational) provides an inroad for the precise study of correspondence between agents interacting with their environments or with each other. Preserving structure of meaning channels for an agent coupled to its environment is required for the usefulness of and determines the quality of mappings in the design, algebraic engineering, interaction dynamics, and constructive biology of groups of situated agents.

1.4. Further Examples from Constructive Biology

Rodney Brooks inspired by the walking of insects has shown how robust behavior, control, and terrain negotiation can be emerge from the interaction of autonomous components (augmented finite-state machines) layered in a way that modulates their interactions (subsumption architecture),[11] and his group at MIT is scaling up these principles in an effort to incrementally construct a humanoid robot called Cog.[12] Kerstin Dautenhahn, working toward grounding social intelligence in multiagent systems, has studied how minimal embodied systems can achieve successful behaviors in robotic scenarios that depend on interaction in such examples as balancing, recognition and recharging, and imitation[13,14,8,15,16] and, together with Aude Billard, acquisition of a proto-language using learning by imitation.[17,18]

Other of the many examples include Bruce Blumberg's ethologically inspired Silas T. Dog, a virtual dog with synthetic emotions and learning which drive sophisticated action-selection mechanisms as the dog interacts with real humans[19]; Auke Jan Ijspeert's digitally evolved virtual lampreys controlled by artificial neural networks[20,21]; the descriptions of genetic control and the biologically realistic artificial retinas evolved to develop in the presence of noise of Hamid Bolouri and collaborators.[22–25]

We have listed just a few of the many good examples in this exciting area, but many challenges remain to be met. Other areas of constructive biology[26] of great active interest include self-maintenance, morphogenesis, evolution of development, evolution and maintenance of individuality and higher units of selection, and cannot be treated here. Here we shall instead concentrate on the most important current global issues for post-reactive robotics and agents.

1.5. Outline of Paper

We proceed by considering the grounding of behavior in homeostasis, the property of biological systems to maintain key aspects of the internal milieu and interaction with the external environment within narrow ranges of important parameters (Section 2) and discuss the role of drives, internal signals (e.g. hormones) in modulating behavior, and connections between emotions (formally defined as state-changes in response to reinforcing stimuli) and learning, in further freeing a biological agent from the lowest levels of blind reactivity. With such mechanisms in mind, meaning and communication for situated agents are addressed. Correspondence between different agent systems (e.g. in imitation) are also briefly treated. In section 4, we discuss methods for widening the temporal scope relevant for control action by means of temporal grounding, and suggest how certain mathematical constructions (semigroups and their expansions) could be employed to approach problems of communication via histories or stories. In the intra-agent case this leads to a notion of *remembering* the agent's own autobiography and experience, and in the inter-agent case to story-telling and the exchange of experience in a social context in which robots and agents can benefit from 'listening to' each other's experiences. We see that the temporal horizon can be broadened along a scale ranging from from merely reactive to post-reactive.

2. GROUNDING: INTERNAL STATES FOR FITNESS AND SURVIVAL

2.1. Emotional Grounding: Taxis, Tropisms, Drives, Motivation

Charles Darwin[27] realized the importance of emotions and their expression in animals, and his lead as been followed by recent builders of building artificial systems.[*] Masanao Toda's Fungus Eaters[28] carry out certain behavioral programs for survival when they have certain "urges"; this results in adaptive behavior. Such urges serve to regulate behavior, yielding appropriate actions in appropriate contexts (e.g. eating when hungry). Rolf Pfeifer[29] discusses how observer-attributed emotions emerge based on the implementation of taxis and drives in simple robotic implementations. Steve Grand[30] has used such notions in implementing CREATURES, a successful product based on artificial life technology in which so-called Norns grow-up in an environment learning and responding to stimuli while being governed by a set of urges (hunger, sex-drive, etc.) that may take on dynamically varying numerical values.

Emotion systems involved in feedback control of situated agents may serve to provide the grounding for embodied agents in the body/environment coupling. Moreover, affect may play an important role in memory and historical grounding. The psychologist R. Zajonc has shown[31] that humans *prefer* stimuli that they have experienced before to new stimuli (*"Familiarity breeds content[ment]"*); whereas humans who have lost some affective capacities due to aphasia are apparently unable to function "rationally" (Damasio[32]).

We now distinguish the feeling of experience (*qualia*) from operational notions of taxis/tropisms, drive, emotion, and relate these concepts to reinforcement and in the next section to learning and meaning.

Taxis / Tropism. Taxes and tropisms are 'hard-wired' approach or avoidance behaviors in response to stimuli, e.g. turning toward light (in plants), or moving up a gradient of food (*E. coli*) or pheromone concentration. The behaviors are stereotyped and not instrumentally arbitrary, i.e. the agent does not employ and cannot even be trained to employ alternative strategies of behavior in response to the stimulus but reacts in a fixed manner.

Drives. Drives are homeostatic and instinctual mechanisms of internal motivational change or modulators of behavior in response to internal aspects of state: hunger, thirst, sex drive, maintaining temperature and other variables within acceptable ranges while interacting to the environment. Needs of self-maintenance and self-production regulated by hormones (blood-borne signals) in the context of internal milieu account for many drives. The appropriate ranges of related parameters maintained in a homeostatic, possibly living, system need not be fixed but may depend dynamically also on cyclical or otherwise varying internal aspects of state; for example this is the case with hunger, and sex drive which varies with history and hormonal state, although either of these at times may be triggered (like an emotion – see below) by external stimuli.

---

[*] This section and the ensuing one closely follow the text of "Constructive Biology of Emotion Systems: First and Second Person Methods for Grounding Adaptive Behavior in a Biological and Social World" by C. L. Nehaniv (submitted for journal publication).

Reinforcing stimuli. Reinforcing stimuli are stimuli which an agent will work to obtain (positive reinforcing stimuli), or to avoid or terminate (negative reinforcing stimuli). In the case of experimental psychology and ethology, this constitutes an operational definition for the identification of stimuli as positively or negatively reinforcing. Stimuli, perceptable to the agent, which for the agent are not reinforcing are called *neutral* or *unconditioned stimuli*; the agent does not seek to avoid or obtain such a stimulus. An unconditioned stimulus need not remain one.

Some stimuli are reinforcing innately, i.e. by the design, nature, or default structuring of the agent, and are referred to as *primary reinforcers*. Proposed categories of primary reinforcers[33] for animals include tastes (salt, sweet, bitter, sour, and others), odor (putreying odors, pheromones), somatosenory (pain, touch, grooming, washing, temperature), certain visual stimuli (symmetry, open blue sky, secure cover), auditory (warning call, vocalizations), reproduction (courtship, mate guarding, nest building, infant attachment to parents, crying of infant, parental attachment), novel stimuli (leading to curiosity), sleep, altruism within kin and social groups, group acceptance, play, and others. According to a first person view, negative reinforcers (punishers) are considered *painful* or *unpleasant* to the agent, whilst positive ones (rewards) are *pleasant* or *enjoyable*.

Unconditioned stimuli which originally had no reinforcement effect (the agent would neither seek to obtain or avoid them) may become associated through Hebbian learning or classical conditioning with reinforcing stimuli. This is called *stimulus-reinforcement association learning*, i.e. the association of a stimulus with an existing reinforcing stimulus. In this way, Pavlov's dog associated the sound of a bell with food.[34] Such learning is to be clearly distinguished from "stimulus-response learning" also called "habit learning", since it is only the *association* of stimuli, and not a response, that is learned. Once a stimulus is associated with a reinforcer, it becomes a *conditioned reinforcer* if the agent will either work to obtain or avoid it. Any reinforcer that is not a primary (unlearned) reinforcer is referred to as a *secondary reinforcing stimulus*.

This picture of primary and secondary reinforcers just painted could be misconstrued as static, but that would be an oversimplification. Stimuli can become or cease to be secondary reinforcers as new associations are learned and old ones forgotten in a changing environment. It can happen that a primary negative reinforcer like pain associated with the eating of a spicy food becomes a secondary positive reinforcer; or in pathological cases animals including humans may work to obtain painful stimulation. Moreover, often the reinforcement value of stimuli dynamically varies with the state of internal parameters and drives. For example, satiety or habituation to a particular stimulus may cause it to lose its reinforcement value temporarily.

Emotions. Emotions are defined as changes in state in response to primary or secondary reinforcing stimuli, or in some cases due the remembering of such stimuli. Notice that since the definition of reinforcer is operational, so is this definition of emotion. The experience of qualia (feeling, awareness, or consciousness) of the state change is a possible but not a necessary aspect of emotion in this formal sense. The operational definition of emotion as state change in response to reinforcing stimuli here follows Gray[35] and Rolls[33] for animals. We observe that it also makes sense for artificial agents. The defining state change may ensue following neural processing, biochemical reactions and physiological changes, motivation and perceptual interplay, rule-based reasoning, cognitive appraisals (Ortony, Collins, and Clore,[36] Roseman *et al.*[37,38]), or in response to changes in bodily configuration and expression (James and Lange[39,40]), combinations of these factors (Izard[41]), or by any means at all. Such types of changes induced by mechanisms listed above have been proposed in various theories and models of human, animal, and agent emotions. For example, Ortony, Collins and Clore[36] define emotions as "valenced reactions to events / agents / objects [...] whose particular nature is determined by the way in which the eliciting situation is construed". Although their cognitive appraisal framework was never intended for the generation of artificial emotion, it has been applied to the this as well as to artificial reasoning about affect in multi-agent systems (Elliot[42]).

In the approaches mentioned above, the type of eliciting stimulus (e.g. object, event, action, or person) and the particular drives related to and sensory characteristics of the reinforcing stimulus contribute to determining the character of the emotion. It is evident that whether the experience is first person or attributed to a second person also contributes in a fundamental way to the character of the emotion.

Studies of human emotion reveal that two dimensions are extremely relevant in what is understood (intuitively rather than formally) to be required for emotion: first, emotions are *valenced*, they are either good or bad, pleasant or unpleasant; and second, they have degree (level of intensity). These properties of emotions imply that they can serve an evaluative function on particular stimuli in the particular situations that result in them. In this way, they can serve as a '*common currency*' (e.g. Rolls[33]) by which to evaluate stimuli and then to *compare* the likely results of various courses of action.

Moods. Moods are longer term changes in system state that persist over extended periods of time and may have strong effects on body and behavior; for example, 'peppiness' may last all morning, while depression may last for years. (Formally, this definition still lacks the rigor of the preceding ones.)

While emotions are state changes in response to stimuli, a mood does not –to use a grammatical metaphor– "have an object", i.e. it is not elicited in relation to a particular object or agent or event in the environment. Any operational distinction between mood and emotion is complicated by the fact that remembering or imagining an environmental stimulus might result in an emotion.

Motivation and intent that arise from remembering or in planning can also guide behavior. Fast reactive responses elicited by some emotions seem to arise through certain limbic neural pathways in animal brains, slower cortical functions and deliberative evaluation may play a role in others, while abstract symbol manipulation may be involved in other highly cognitive emotions and possibly in consciousness. There may be several pathways to action, mediated by several levels at which drives and emotions arise and are arbitrated amongst. A three-layer (reactive, deliberative, and self-monitoring) architecture proposed by Aaron Sloman and collaborators as a model of human-like emotion for example realizes such division of labor in the control of behavior.[43,44]

3. EMOTION IN ADAPTIVE SYSTEMS

3.1. Time and Emotion

Temporal aspects of the stimuli-reinforcer association, along with the type of stimulus, are extremely important in the class of emotion elicited. That is, the temporal extent of the learned stimulus may precede, coincide with, follow, or overlap in several possible ways the duration of the experience of the reinforcer. For example, the sound of a bell may be predictive of food if it has always preceded food; hearing it may lead to emotions of expectation [anticipation, hope – or in a negative case, dread and fear], or, if the positive [resp. negative] reinforcer is not forthcoming, to disappointment [resp. relief]. Alternatively, if the positive [resp. negative] reinforcing stimulus does occur then 'hopes confirmed (satisfaction)' [resp. 'fear confirmed'] state changes define the resulting emotion. First-person actions *preceding* reinforcing stimuli can elicit such emotions as guilt, regret, shame, pride; or pleasant surprise (unexpected reward), unpleasant surprise (unexpected punisher), or neutral surprise (unexpected nonreinforcing stimulus – formally, this last is a state change in response to new information but is *not* an emotion since it does not involve a reinforcer unless perhaps the *novelty* itself is reinforcing). Some factors contributing to the character of complex (social) emotions such pride, guilt or shame are the attribution of observer status to others who may perceive the first-person's action.

Varying temporal configuration, the results in state change of the agent might be somewhat different if the bell had always co-occurred with food, e.g. disappointment might be more immediate if no food were presented during the sound of the bell than if food had always been presented only sometime after or before the sound of the bell. The two later conditions could first result respectively in positive anticipation and confusion as immediate effects rather than disappointment. Thus the relative temporal configuration of associated stimuli influences the character of emotion. One artificial neural network architecture that can learn associations together with the relative temporal configuration is DRAMA, developed by Aude Billard.[45,46] We return to the consideration of temporal aspects and emotion in section 4 below when we discuss the temporal horizon of humans and other animals, as well as the possible implications for constructed agents, in response to ideas of Heidegger.

3.2. Emotion and Learning

Since emotions are changes in state elicited by reinforcing stimuli, their valence and degree can serve as measures of the [un]desirability of pursuing a course of action that leads to further stimuli. In particular, the particular course of action to take in obtaining or avoiding the same or an associated stimulus is not encoded in either the valence or degree of the emotion, yet the agent can take this valence and degree as a guide to suggest a course of action: to work (somehow) either to obtain or to work to avoid or terminate a stimulus.

How the agent works to obtain a stimulus can be to choose to invoke more general strategies and behaviors generically applicable to large classes of situations: e.g. approach, grab, flee, hide. In this way, stimulus and response are de-coupled and the relations for behavior in response to a stimulus are modifiable, dynamically learnable and reconfigurable. Thus the common currency of emotion can serve to modulate the control of the agent and to motivate or suppress certain responses in its interaction with the world.

This provides a mechanism for a *"two-process theory of learning"* (Mowrer,[47] Gray,[35] Rolls[33]). This type of effect of synthesized emotion on learning can be found implicitly for instance Bruce Blumberg's Silas T. Dog, a synthetic worlds virtual agent, which attempts to determine which aspects of input (external stimuli) elicited an internal state change (e.g. increase in a 'fear variable') and learns the association of the stimulus with the formal emotional change. The latter association influences, in a rudimentary way, the agent's learning and behavior (e.g. avoidance of locations where an unpleasant stimulus was encountered).[48]

This two-process model is distinct from the behaviorist's operational analyzed model of stimulus-response learning (Hull,[49] Spence,[50] Skinner[10]), which it factors into (1) stimulus-reinforcer association (see above) and (2) the learning of responses. In contrast to the learning of fixed stimulus-response pairs, this factorization allows an approach for flexibility of behavioral responses to reinforcing stimuli and the capacity to substitute one behavior for another if the first fails to achieve the desired effect. It is useful as part of a constructive biology approach addressing the possible internal mechanisms relating affect, learning and behavior.

Moreover, the *expression* of an emotional or drive state may be perceptable to conspecifics or other agents (prey or predators). Recognition of the expression of the other can serve as a index to its state: e.g. the other's recent experience of reinforcing stimuli (signs of seeing a tasty victim, instinctual alarm calls), and hence as an indicator of its intent (that it might work to obtain or avoid something in the environment). Certainly in the case of biological evolution, it could be adaptive to use the information expressed in such signals to avoid predators, assess the state of prey, or gauge the likely behavior of conspecifics. Such use is a second-person method of adaptation.

Possible design approaches to making responses that exploit these signals include: (1) to rely on natural selection and evolution (over generations) or (2) to rely on learning and adaptation (within an individual). The systematicity of either (1) or (2) in associations of signal-from-others with appropriate behavioral response can be *ad hoc*, partial, or comprehensive, and could vary in degree of flexibility.

A comprehensive or partial correspondence from signals perceived in the other agent and how to respond with one's own action, can be achieved by incremental learning, or it could take advantage of a natural correspondence – the identification of the other agent as a 'second person' with a similar architecture to one's own, at least similar enough that some prediction could be accurately made of the state of the agent based on the signals generated, together with prediction of likely action that one would make in such a state ('mind reading' or 'reading of intent'[51]). We hypothesize that socially intelligent animals species make widespread use of such systematic second-person mechanisms. (See (Dautenhahn[16]) for related issues of social intelligence in animal species with individualized societies.)

Strategies (1) and (2), with varying systematicity and flexibility, may be applied to the design or the explanation of the interaction with others of a particular first-person agent.

Further aspects of second person mapping are addressed in the general study of imitation among differently embodied systems.[52] Solving the *correspondence problem* — or *trying to imitate* — between the body of another (or the perceptions of one's own body) and what possible actions one could make that would correspond to what is perceived.[13,53] Solutions to this problem are not unique and depend on observer criteria as well as on the particular details of embodiment. Once the correspondence is solved or partially solved (possibly by design), *learning by imitation* is possible, i.e. the robot, animal or artifact can use successful imitation as a scaffolding to acquire new behavioral competences.[54]

3.3. Meaning and Communication

Meaning is understood here as *(1) information in interaction games between an agent and its environment or between agents mediated with respect to their own sensors and actuators* and as *(2)* useful *for satisfying homeostatic and other drives, needs, goals or intentions.*[55]

Meaningfulness may be in the designer's eye or in the adaptiveness of the activity as tending to increase the probability that copies of an agent's genes (if it has any) are represented in future generations. The latter notion of evolutionary, behavioral, survival adaptiveness (in biological agents the tendency to increase reproductive success) hints at the possible nature of meaning for evolved or constructed systems. *Meaning arises with information that helps an agent attain 'goals'.* However, the goals may or may not come from the agent itself, but be attributed by an external observer or may serve some higher level unit of selection to which the agent belongs. For example, if robots interact to perform a task such as collecting objects (Beckers *et al.*[56]), meaning arising in meeting the external observer's goal (collecting) rather any robot's goal. In the analogous collective behavior by termites,[57] particular local actions achieve only individual termites' local goals, from whose combined effects emerge colony-level goals (such as heat-regulation in the termite mound), which are adaptive for a higher level unit of selection (the colony as agent).

Note that meaning in this sense starkly contrasts to – but may also be considered a compatible refinement of – Shannon's measure of information content,[58] which is minimal for a constant unchanging signal but is maximal for random signals, both of which might well be devoid of meaning for all agents and observers. Agent goals may be conscious or unconscious, merely surviving, reacting, maintaining the self or reproducing, or they may motivate actions according to intentionality. In fact it may be that the agent has no proper goals but that its actions reflect only a merely reactive coupling to the environment. If the goals are observer-attributed rather than within the agent then the corresponding meaning exists only in relation to such observers. The agent itself may be such an observer, in which case meaning could then arise for it in its interaction with its *Umwelt* ("world around").

Meaning then need not be linguistically nor even symbolically mediated. It may or may not involve representations, but must arise in the dynamics realizing the agent's functioning and interaction in its environment (cf. the notion of 'structural coupling' of Maturana and Varela[59]), supporting adaptive or self-maintaining or reproductive behaviors, or goals, or possibly intentions or plans. *Multiple observers*, as in the case of interaction among human agent/observers, *result in multiple arisings of meaning.* Any entity that exists at the level of a biological unit of evolutionary selection (e.g., unicellular organism, differentiated multicellular organism, eusocial insect colony) could potentially be an agent or observer in our sense, as could a human organization. Of course, robotic and software agents are not excluded.

In the realm of constructive biology, robotics and artificial agent construction, meaning can also arise in the interaction channels between the agent and the environment in which it is 'embodied'. These channels could be artificially evolved or designed. Similarly, these considerations apply to software agents, which might in a sense be considered embodied with respect to their particular environments as long as mutually perturbing channels exist between the agent and its environment (this ontology-independent definition of *embodiment* is due to Tom Quick[60]), with *degree of embodiment* measurable according to the complexity of the dynamics occurring between the two.

The philosopher Ludwig Wittgenstein (in his later, important work) insisted on defining meaning of words and other signs in terms of their *use* by agents engaged in language games (including artificial and everyday language).[61,62] An insight going back to the 19th century of C. S. Peirce,[63,64] the father of semiotics, is that signs *mediate* meaning, only make sense in the context of systems of signs, and that an *interpretant* always links a signifier to a signified in an embedded and embodied process (*semiosis*). This situated and embodied nature of agent semiotics highlights the meaninglessness of signals, signs, and sign systems in isolation, without agents and thus without uses. Signal and sign systems may or may not have formally specifiable structures. They may be difficult to describe, prescribe or construct for given competences and desired performances in various *interaction games.*

We note that there is no fundamental reason to restrict Wittgenstein's insights to *language games* or the 'language' of interaction games to verbal utterances. Other kinds of signals and actions can also be used by an agent interacting with its environment. Thus we speak of *interaction games* as a generalization of Wittgenstein's language games. The partner in an interaction game may be another agent, or it may be the environment where the agent is situated. The agent interacts in the game by accessing channels of meaning.

4. TEMPORAL GROUNDING

A feature of memory and remembering is that they provides 'extrasensory' meaningful information by which an agent may modulate or guide its immediate or future behavior. With small temporal scope, this also occurs with moods and emotions. Remembering involves simple or complex narrative structure, and so communication of narrative (state-histories, memories, or stories) between agents can provide 'extrasensory' channels of meaning.

Martin Heidegger[65] saw the state of man as being as situated in the Now, being here in the imminence of the Future in relation to the impinging Past. This *temporal horizon* is extremely broad in humans compared to other animals, as is evidenced by our emotions such as hope and regret, concern with planning for future actions and story-telling about past or imagined events. This vast temporal horizon means that humans will tend to deal with interaction in a way that makes narrative sense, and may anthropomorphically expect their technological agents to do so. Affect and narrativity thus intertwine with each other. Extrasensory data from narrative and historical temporal grounding helps an agent escape from the present in its perception-action cycle.

The cost and reward of experience stimuli provide a uniform dimension in which to evaluate the result or desirability of action, and the relative values of these costs and rewards (or 'pains' and 'pleasures') may be modulated by current state of the agent. Most attempts to introduce 'emotion-parameters' into AI systems, can be seen as an attempt to solve the well-known contextualization problem in AI and robotics, i.e. to transcend simple reactivity by allowing the settings of the parameters to modulate behavior, so as to respond appropriately to the given context.

A way to go further than the use of emotion synthesis in temporal grounding is via the use of the methods for narrative and historical grounding described below.[†] This outlines the programme of applying algebraic methods in story-telling first introduced by C. L. Nehaniv and K. Dautenhahn.[66,9]

4.1. Life, Time, and History

Embodied biological and artificial agents have histories (usually irreversibly) reflected in their structure. Without the historical context (phylogeny and ontogeny) we cannot understand the structure, appearance and behavior of biological agents. In formal terms we can say that such historically embedded agents are subject to algebras of events. Memory of the histories and stories which these agents undergo and create during their life times can also be of use in synthesizing agents that need to act in socially intelligent ways, possibly in cooperative tasks or so that they appear believable and acceptable to humans. The concepts *autobiographic agent* and *biographic reconstruction* (sec. 1.3 above) are useful in achieving a broad temporal horizon for post-reactive systems. We now discuss the grounding of algebras of time and of history to support the re-construction of these agents' own histories as well as the stories of other agents.

Each natural living system can only be understood best and its phenotype can only be interpreted best by taking into account its historical context, its evolutionary context, phylogeny, and individual developmental context, ontogeny. Each creature has its own history. Biological systems exist under circumstances of irreversibility that make them fundamentally different from inanimate matter (distinguishing them for instance as subject to development, damage and death). Individual historical memory and story-telling are capacities useful in coping with this irreversibility. Such mechanisms may for embodied agents contribute to — or may be even be necessary for — competence in social intelligence. Thus, it seems worth studying how behavioral complexity emerges from a story-telling/autobiography reconstructing system acting and interacting in its environment.

4.2. Algebraic Tools

Structuring historical memories connects to an area of algebra called global semigroup theory, which allows one to construct expressions in 'algebras of time' that can support recording events of fundamental significance to an agent. This may include records of internal changes (e.g. in motivational, goal, emotional, cognitive, body image, perceptual states, knowledge of other entities) as well as external changes (location, objects or other agents in the environment, physical conditions of local environment) that provide grounding for the representation. The models of change and time selected can be different depending on what is important to the agent.

Time has the property that a sequence of events c following the two successive event sequences a then b is the same as b followed by c and prior to both a. Thus sequences of events in time satisfy the *associative law*: $(ab)c = a(bc)$. Structures satisfying this law are called *semigroups*. Unlike the time of differential equations (which fails to be expressive enough for living systems), associativity gives us a framework for myriad models of time, even if irreversible. For any agent, the sequences of events it may experience and the manner in which these transform its state determine a semigroup structure, which is the *algebra of time* for that agent.

Expansions are systematic treatments of histories in algebras of time.[‡] Recording histories of various kinds can be used to systematically 'expand' algebras of time. This corresponds to an expansion of the temporal horizon – by using an expansion rather than the original algebra of time, it is possible to express formal stories or histories.

[†] The remainder of this section closely follows some of the text of "Embodiment and Memories — Algebras of Time and History for Autobiographic Agents" by C. L. Nehaniv and K. Dautenhahn from *Cybernetics and Systems'98*, Proceedings of the 14th European Meeting on Cybernetics and Systems Research, edited by Robert Trappl, published by the Austrian Society for the Cybernetic Studies, Vol. 2, pp. 651–656, 1998.

[‡] Formally, an *expansion* $F : Sgp \rightarrow Sgp$ is a functor from the category of semigroups to itself and a natural transformation η from F to the identity functor, associating a surjective homomorphism $\eta_S : F(S) \rightarrow S$ to each semigroup S. Thus $F(S)$ may be regarded as 'blow-up' or expansion of S.

The choice of what to remember in the history is by no means unique. Some of the myriad examples for how to select what to remember include: *Newtonian histories* consist of records of all instances of physical time and associate events that occurred at each. Such a notion of historical time is somewhat bizarre from the perspective of living systems for which only especially important events may be relevant, and so ultimately probably it will not be of great interest for them or for use in other embodied agents. This realization about individual relevance of events is closer to Freud's notion of intense or traumatic events as determinants of an individual's character (birth, first sex, marriage, loss of parent, ...), which comprise the elements of a *Freudian history.*[§] Other notions of history may include records of reflections on the past events or future events (applicable in planning, but also in neurosis), and so on. These different approaches to histories all correspond to different semigroup expansions in algebra.

The semigroup for possible events in the life-history of a robot or organism varies strongly with the structure of the entity and its sensory, perceptual and other capacities. Given a semigroup, expansions are uniform ways of enlarging the semigroup into another one whose elements include formulaic encodings of historical information about events possible in the semigroup. It is thus possible to expand various heterogeneous semigroup structures in the same way, e.g. Newtonian histories can be constructed for two physically radically divergent entities. Here the fact of designed, evolved or adapting interfaces to the world or to other entities or to portions of the self, which *could have been otherwise*, indicates the need for a mathematical language that can usefully formalize the notion of 'what happened'.

The elements of expansions themselves can be combined using an associative multiplication. Recording them, augmenting them, changing them, and transmitting them are formal analogues for remembering and story-telling. Understanding the story of another agent may make sense in light of mappings that show how its states and experiences correspond to one's own (i.e. homomorphisms — structure-preserving maps – allow one to relate meaning for the first and second person).

4.3. Story-Telling and Memories

The problem of communicating knowledge of other minds could be approached in this framework by the passing of algebraic expressions (in a historical 'expanded' semigroup) which record changes in the life-long processes of learning and experiencing. Such expressions ('histories') may refer to the agent's own experiences or the experiences of another agent. Passing such an expression would correspond to revealing one's autobiography or, more generally, to telling a story. We consider story-telling as a basic element of remembering and re-construction of experiences of one agent, and as a central element in social communication, dialogues. Such autobiographies could of course be communicated to a human or artifact social agent. Furthermore, the algebraic structure on histories means that one has algorithms to build-up computed histories from successive portions of history. That is, elements of a formal algebraic story may be multiplied (combined) by a receiving agent to yield a history of the sending agent. Such an approach may provide insight into the representation by social agents of other minds, which has been suggested as a fundamental pre-requisite for social intelligence.

Roger C. Schank and Robert P. Abelson[67] discuss that mechanisms of remembering, perceiving and re-interpreting the world – in particular the social world – are mainly based on 'stories'. Stories should be fundamental to human memory, knowledge, and social communication. Experiences are interpreted in terms of old stories, new experiences are understood by retrieving and re-creating a similar story which is consistent to the new story, using the embodied 'self' as the point of reference. Understanding depends on whether a good 'old' story can be recreated by the input of a new story. The more stories an agent can tell, the greater its potential of understanding and social communication, since communication can be thought of as mutual story-telling behavior. The most effective means to communicate an experience or knowledge is to put it into a good story. A pleasant dialogue is given when two agents have the feeling of 'being understood' (which is not necessarily identical to sharing exactly the same opinions and attitudes), namely when one agent is telling appropriate stories as a response to the story the other agent has told. Stories and therefore memories change in the process of being told and remembered; sharpening and levelling processes lead to modifications of content and meaning of stories. Remembering a story (maintaining it as part of one's autobiography) requires telling it, with others as audience and interpreter of our stories, even if perhaps these 'others' are only versions of the self. Thus, story-telling can only be meaningful (useful) in temporally grounded and

---

[§] We emphasize that we use the word 'Freudian' only in an extremely limited sense: Freud's deep insight for psychology was that one should look at important events in an individual's past. Our use of the term should *not* be construed as acceptance of Freud's models of human cognition or cognitive development, but has a strict mathematical sense.

social context. According to Read and Miller,[68] social context may have provided the reason why human cognition developed toward story-telling, since stories are simply very effective re-presentations to code complex experience.

Giving artificial agents such a capacity could result in historically embedded artifacts which can tell ‘good stories’ to other agents as well as to humans. It could also release them from the lowest levels of reactivity.

Worlds of experience have temporal horizons limited in various ways:
(1) reactive systems – nearly completely limited to the *now*, with only minimal impact of internal state;
(2) affective systems – systems whose drives, motivations, and emotions (as formally defined above) modulate their attainment of the goals of fitness and survival and help contextualize behavior;
(3) learning affective systems – systems like those of type (2) which employ learning (e.g. Q-learning, associative DRAMA-type learning, or two-process learning) – of course, learning without affect can and has been added to systems of type (1);
(4) post-reactive temporally grounded systems acting with respect to a broad temporal horizon, story-telling and remembering systems, autobiographic agents, systems with higher various degrees of social and narrative intelligence.

Post-reactive robotics can be approached using algebras of time (semigroups) and their expansions (algebras of history) to capture the temporal nature of being. Histories, memories, and shared stories must be grounded in the interaction games and channels of meaning of the particular agent. In a constructive biology of post-reactive systems, the sky over a broad temporal horizon opens for action-selection to take better advantage of prior experience, anticipation, and future goals of the self and others, remembered or communicated in stories.

REFERENCES

1. J. Von Neumann, *Theory of Self-Reproducing Automata, (Arthur W. Burks, ed.)*, University of Illinois Press, Urbana, Illinois, 1966.
2. C. G. Langton, “Self-reproduction and cellular automata,” *Physica D* **10**, pp. 135–144, 1984.
3. J. D. Watson and F. H. C. Crick, “A structure for deoxyribose nucleic acid,” *Nature* **171**, pp. 737–738, 1953.
4. B. Webb, “Robotic experiments in cricket phonotaxis,” in *From Animals to Animats 3: Proceedings of the Third International Conference on Simulation of Adaptive Behavior (August 8-12, 1994, Brighton, England)*, D. Cliff, P. Husbands, J.-A. Meyer, and S. W. Wilson, eds., pp. 45–54, 1994.
5. B. Webb, “Using robots to model animals: A cricket test,” *Robotics and Autonomous Systems* **16**, pp. 117–134, 1995.
6. V. Braitenberg, *Vehicles: Experiments in Synthetic Psychology*, MIT Press, 1986.
7. C. L. Nehaniv, “The second person - meaning and metaphor,” in *Computation for Metaphors, Analogy, and Agents*, vol. 1562, Lecture Notes in Artificial Intelligence, pp. 380–388, Springer Verlag, 1999.
8. K. Dautenhahn, “I could be you - the phenomenological dimension of social understanding,” *Cybernetics and Systems Journal, Special Issue on Epistemological Aspects of Embodied AI* **28(5)**, pp. 417–453, 1997.
9. C. L. Nehaniv and K. Dautenhahn, “Embodiment and memories — algebras of time and history for autobiographic agents,” in *Cybernetics and Systems ’98 : Proc. 14th European Meeting on Cybernetics and Systems Research (Vienna, Austria – 14-17 April 1998)*, R. Trappl, ed., vol. 2, pp. 651–656, Austrian Society for Cybernetic Studies, 1998.
10. B. F. Skinner, *The Behavior of Organisms*, Appleton-Century, New York, 1938.
11. R. A. Brooks, “A robust layered control system for a mobile robot,” *IEEE Journal Robotics and Automation* **RA-2**, pp. 14–23, April 1986.
12. R. A. Brooks, C. Breazeal, M. Marjanović, B. Scassellati, and M. M. Williamson, “The Cog project: Building a humanoid robot,” in *Computation for Metaphors, Analogy, and Agents*, C. L. Nehaniv, ed., vol. 1562, Lecture Notes in Artificial Intelligence, pp. 52–87, Springer Verlag, 1999.
13. K. Dautenhahn, “Trying to imitate - a step towards releasing robots from social isolation,” in *Proc. From Perception to Action Conference (Lausanne, Switzerland)*, P. Gaussier and J.-D. Nicoud, eds., pp. 290–301, IEEE Computer Society Press, 1994.
14. K. Dautenhahn, “Getting to know each other - artificial social intelligence for autonomous robots,” *Robotics and Autonomous Systems* **16**, pp. 333–356, 1995.
15. K. Dautenhahn, “Biologically inspired robotic experiments on interaction and dynamic agent-environment couplings,” in *Proc. Workshop SOAVE’97 Selbstorganization von Adaptivem Verhalten (Ilmenau – 23-24 September 1997)*, pp. 14–24, 1997.

16. K. Dautenhahn, "Embodiment and interaction in socially intelligent life-like agents," in *Computation for Metaphors, Analogy, and Agents*, C. L. Nehaniv, ed., vol. 1562, Lecture Notes in Artificial Intelligence, pp. 102–142, Springer Verlag, 1999.
17. A. Billard and K. Dautenhahn, "Grounding communication in autonomous robots: an experimental study," *Robotics and Autonomous Systems* **24**(1-2), pp. 71–79, 1998.
18. A. Billard and K. Dautenhahn, "Experiments in social robotics: grounding and use of communication in autonomous agents," *Adaptive Behavior* **7**(3), to appear 1999.
19. B. Blumberg, *Old Tricks, New Dogs: Ethology and Interactive Creatures.* PhD thesis, MIT Media Lab, 1996.
20. A. J. Ijspeert and K. Kodjabachian, "Evolution and development of a central pattern generator for the swimming of a lamprey," *Artificial Life* , to appear 1999.
21. A. J. Ijspeert, J. Hallam, and D. Willshaw, "Evolving swmming controllers for a simlated lamprey with inspiration from neurobiology," *Adaptive Behavior* **7(2)**, to appear 1999.
22. C. Yuh, H. Bolouri, and E. H. Davidson, "Genomic cis-regulatory logic: Experimental and computational analysis of a sea urchin gene," *Science* **279**(5358), pp. 1896–1902, 20 March 1998.
23. H. Bolouri, R. Adams, S. George, and A. G. Rust, "Molecular self-organisation in a developmental model for the evolution of large-scale neural networks," in *International Conference on Neural Information Processing and Intelligent Information Systems (ICONIP'98)*, vol. 2, pp. 797–800, 1998.
24. A. G. Rust, S. George, and H. Bolouri, "Developmental evolution of an edge detecting retina," in *Proc. 8th International Conference on Artificial Neural Networks (ICANN'98)*, L. Niklasson, M. Boden, and T. Ziemke, eds., vol. 2, pp. 561–566, 1998.
25. M. J. Loomes, R. Frank, and N. Davey, "Theory building - a constructive approach to modelling neural network development," in *Selected papers from Neural Networks - Producing Dependable Systems*, vol. 2.1.1-2.1.10, 1997.
26. C. L. Nehaniv, ed., *Mathematical and Computational Biology: Computational Morphogenesis, Hierarchical Complexity, and Digital Evolution*, vol. 26, Lectures on Mathematics in the Life Sciences, American Mathematical Society, Providence, Rhode Island, 1999.
27. C. Darwin, *The Expression of Emotion in Man and Animals*, 1892. Reprinted by the University of Chicago Press, 1965.
28. M. Toda, *Man, Robot, and Society*, Nijhoff, The Hague, 1982.
29. R. Pfeifer, "The fungus eater approach to emotion: a view from artificial intelligence," *Cognitive Studies* **1**, pp. 42–57, 1994.
30. S. Grand, D. Cliff, and A. Malhorta, "CREATURES: Artificial life autonomous agents for home entertainment," in *Proc. First International Conference on Autonomous Agents (AGENTS'97 – Marina del Rey)*, Association for Computing Machinery, February 1997.
31. R. B. Zajonc, "Feeling and thinking: Preferences need no inferences," *American Psychologist* **35**, pp. 151–175, 1980.
32. A. R. Damasio, *Descartes' Error: Emotion, Reason and the Human Brain*, G. P. Putnam & Sons, New York, 1994.
33. E. T. Rolls, *The Brain and Emotion*, Oxford, 1999.
34. I. P. Pavlov, *Lectures on Conditioned Reflexes,* W. H. Gantt, trans., Liverwright, New York, 1928.
35. J. A. Gray, *Elements of a Two-Process Theory of Learning*, Academic Press, 1975.
36. A. Ortony, G. L. Clore, and A. Collins, *The Cognitive Structure of Emotions*, Cambridge University Press, 1988.
37. I. J. Roseman, "Appraisal determinants of discrete emotions," *Cognition and Emotion* **5**(3), pp. 161–200, 1991.
38. I. J. Roseman, A. A. Antoniou, and P. E. Jose, "Appraisal determinants of emotions: Constructing a more accurate and comprehensive theory," *Cognition and Emotion* **10**(3), pp. 241–277, 1996.
39. W. James, "What is emotion?," *Mind* **9**, pp. 118–205, 1884.
40. W. James and C. G. Lange, *The Emotions*, 1922.
41. C. E. Izard, "Four systems for emotion activation: Cognitive and noncognitive processes," *Psychological Review* **100**(1), pp. 68–90, 1993.
42. C. D. Elliot, *The Affective Reasoner: A Process Model of Emotions in a Multi-Agent System.* PhD thesis, Department of Computer Science, Northwestern University, 1992.
43. A. Sloman and M. Croucher, "Why robots will have emotions," in *Proc. Seventh International Conference on AI*, pp. 197–202, August 1981.

44. L. Wright, A. Sloman, and L. Beaudoin, "Towards a design-based analysis of emotional episodes," *Philosophy, Psychiatry and Psychology* **3**(2), pp. 101–126, 1996.
45. A. Billard and G. Hayes, "DRAMA, a connectionist architecture for control and learning in autonomous robots," *Adaptive Behavior* **7**(1), January 1999.
46. A. Billard, DRAMA*, a Connectionist Model for Robot Learning: Experiments in Grounding Communication through Imitation in Autonomous Robots*. PhD thesis, Department of Artificial Intelligence, University of Edinburgh, 1999.
47. O. H. Mowrer, *Learning Theory and Behavior*, Wiley, 1960.
48. R. Picard, *Affective Computing*, MIT Press, 1997.
49. C. L. Hull, *Principles of Behavior*, Appleton Century Crofts, New York, 1943.
50. K. W. Spence, *Behavior Theory and Conditioning*, Yale University Press, 1956.
51. R. W. Byrne and A. Whiten, *Machiavellian Intelligence*, Clarendon Press, 1988.
52. K. Dautenhahn and C. L. Nehaniv, eds., *Proc. AISB'99 Symposium on Imitation in Animals and Artifacts, (Edinburgh, Scotland - April 6-9, 1999)*, Society for the Study of Artificial Intelligence and Simulation of Behaviour, 1999.
53. C. L. Nehaniv and K. Dautenhahn, "Of hummingbirds and helicopters: An algebraic framework for interdisciplinary studies of imitation and its applications," in *Learning Robots: An Interdisciplinary Approach*, J. Demiris and A. Birk, eds., World Scientific Press, in press.
54. K. Dautenhahn and C. L. Nehaniv, eds., *Imitation in Animals and Artifacts*, MIT Press, forthcoming.
55. C. L. Nehaniv, "Meaning for observers and agents," in *Proc.* IEEE *International Symposium on Intelligent Control/Intelligent Systems and Semiotics, ISIC/ISAS'99 (15-17 September 1999 – Cambridge, Massachusetts)*, p. (in press), 1999.
56. R. Beckers, O. E. Holland, and J. L. Deneubourg, "From local actions to global tasks," in *Artificial Life IV*, R. A. Brooks and P. Maes, eds., pp. 181–189, MIT Press, 1994.
57. P. P. Grassé, "La reconstruction du nid et les coordinations inter-individuelles chez *bellicositermes natalensis* et *cubitermes* sp. la theorie de la stigmergie: Essai d'interpretation des termites constructeurs," *Ins. Soc.* **6**, pp. 41–48, 1959.
58. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, 1963.
59. H. R. Maturana and F. J. Varela, *The Tree of Knowledge: the Biological Roots of Human Understanding, revised edition*, Shambala Publications, Inc., 1992.
60. T. Quick, K. Dautenhahn, C. Nehaniv, and G. Roberts, "On bots and bacteria: Ontology-independent embodiment," in *Proc. ECAL'99, Fifth European Conference on Artificial Life (Switzerland - September 1999)*, (in press).
61. L. Wittgenstein, *The Blue and Brown Books*, Harper & Brothers, 1958.
62. L. Wittgenstein, *Philosophical Investigations, (Philosophische Untersuchungen), German with English* translation by G. E. M. Anscombe, 1964, Basil Blackwell, Oxford, reprinted 3rd edition, 1968.
63. C. S. Peirce, *Collected Papers*, vol. 2: Elements of Logic, Harvard, 1965.
64. J. Goguen, "An introduction to algebraic semiotics, with application to user interface design," in *Computation for Metaphors, Analogy, and Agents*, C. L. Nehaniv, ed., vol. 1562, Lecture Notes in Artificial Intelligence, pp. 242–291, Springer Verlag, 1999.
65. M. Heidegger, *On Time and Being*, Harper Torchlight Books, 1972.
66. C. L. Nehaniv and K. Dautenhahn, "Semigroup expansions for autobiographic agents," in *Proceedings of the First Symposium on Algebra, Languages, and Computation (Aizu, Japan – 30 October-1 November 1997)*, T. Imaoka and C. Nehaniv, eds., pp. 77–84, Osaka University, 1998.
67. R. C. Schank and R. P. Abelson, "Knowledge and memory: The real story," in *Knowledge and Memory: The Real Story*, R. S. Wyer Jr., ed., pp. 1–85, Lawrence Erlbaum Associates, 1995.
68. S. Read and L. C. Miller, "Stories are fundamental to meaning and memory: for social creatures could it be otherwise?," in *Knowledge and Memory: The Real Story*, R. S. Wyer Jr., ed., pp. 139–152, Lawrence Erlbaum Associates, 1995.